SEQUENCE LISTING

(1)	,	GENERAL	INFORMATION:
-----	---	---------	--------------

- (i) APPLICANT: LI, ET AL.
- (ii) TITLE OF INVENTION: Neurotransmitter Transporter
- (iii) NUMBER OF SEQUENCES: 2
 - (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI, STEWART & OLSTEIN
 - (B) STREET: 6 BECKER FARM ROAD
 - (C) CITY: ROSELAND
 - (D) STATE: NEW JERSEY
 - (E) COUNTRY: USA
 - (F) ZIP: 07068
 - (V) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: 3.5 INCH DISKETTE
 - (B) COMPUTER: IBM PS/2
 - (C) OPERATING SYSTEM: MS-DOS
 - (D) SOFTWARE: WORD PERFECT 5.1
 - (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE: SUBMITTED HEREWITH
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: FERRARO, GREGORY D.
- (B) REGISTRATION NUMBER: 36,134
- (C) REFERENCE/DOCKET NUMBER: 325800-118
- TELECOMMUNICATION INFORMATION: (ix)
 - (A) TELEPHONE: 201-994-1700

(B) TELEFAX:

201-994-1744

- (2) INFORMATION FOR SEQ ID NO:1:
 - SEQUENCE CHARACTERISTICS (i)
 - (A) LENGTH: 2,486 BASE PAIRS
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: CDNA
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGGAGGCAGG	GAGTGAGGAG	CGAGCGGAGT	CGCGTGCGCC	GGCGCGAGCT	CCGGGTCGCC	60
CCAGCCCCAG	CCGGGGGCCT	GTGGCGGGG	AGGAGCTGTG	CGTCCGCGAC	CCGTCGGGA	120
TCGCAGCTGC	TCGGCCGGAG	TGCACGGGCC	GAGTCTGCGC	GACTACCCAC	GCGTGACAGG	180
TCCCTGAATG	AGAAGGAGCT	GACAGCAGCT	GAATTCCATC	TTCTCTGTGT	GCTGGGGAGC	240
AGGGCTACAC	GGCCCAGGTG	GCATCAATGC	CGAAGAACAG	CAAAGTGACC	CAGCGTGAGC	300
ACAGCAGTGA	GCATGTCACT	GAGTCCGTGG	CCGACCTGCT	GGCCCTCGAG	GAGCCTGTGG	360
ACTATAAGCA	GAGTGTACTG	AATGTGGCTG	GTGAGGCAGG	CGGCAAGCAG	AAGGCGGTGG	420
AGGAGGAGCT	GGATGCAGAG	GACCGGCCGG	CCTGGAACAG	TAAGCTGCAG	TACATCCTGG	480
CCCAGATTGG	CTTCTCTGTG	GGCCTCGGCA	ACATCTGGAG	GTTCCCCTAC	CTGTGCCAGA	540
AAAATGGAGG	AGGTGCTTAC	CTGGTGCCCT	ACCTGGTGCT	GCTGATCATC	ATCGGGATCC	600
CCCTCTTCTT	CCTGGAGCTG	GCTGTGGGTC	AGAGGATCCG	CCGCGGAAGC	ATCGGTGTGT	660
GGCACTATAT	ATGTCCCCGC	CTGGGGGGGA	TCGGCTTCTC	CAGCTGCATA	GTCTGTCTCT	720
TTGTGGGGCT	GTATTATAAT	GTGATCATCG	GGTGGAGCAT	CITCTATITC	TTCAAGTCCT	780
TCCAGTACCC	GCTGCCCTGG	AGTGAATGTC	CTGTCGTCAG	GAATGGGAGC	GTCGCAGTGG	840
TGGAGGCAGA	GTGTGAAAAG	AGCTCAGCCA	CTACCTACTT	CTGGTACCGA	GAGGCTTTGG	900
ACATCTCTGA	CTCCATCTCG	GAGAGTGGGG	GCCTCAACTG	GAAGATGACC	CTGTGCCTCC	960
TCGTGGTCTG	GAGCATCGGG	GGGATGGCTG	TCGGTAAGGG	CATCCAGTCC	TCGGGGAAGG	1020

TGATGTATTT CAGCTCCCTC TTCCCCTACG TGGTGCTGGC CTGCTTCCTG GTCCGGGGGT 1080 TGTTGTTGCG AGGGGCAGTT GATGGCATCC TACACATGTT CACTCCCAAG CTGGTCAAGA 1140 TGCTGGACCC CCAGGTGTGG CGGGAGGTAG CTACCCAGGT CTTCTTTGGC TTGGGTCTGG 1200 GCTTTGGTGG TGTCATTGTC TTCTCCAGTT ACAATAAGCA GGACAACAAC TGCCACTTCG 1260 ATGGCGCCCT GGTGTCCTTC ATCAACTTCT TCACGTCAGT GTTGGCCACC CTCGTGGTGT 1320 TTGTTGTTTT GGGCTTCAAG GCCAACATCA TGAATGAGAA GTGTGTGGTC GAGAATGCTG 1380 AGAAAATCCT AGGGTACCTT AACACCAACG TCCTGAGCCG GGACCTCATC CCACCCCACG 1440 TCAACTTCTC CCACCTGACC ACAAAGGACT ACATGGAGAT GGACAATGTC ATCATGACCG 1500 TGAAGGAGGA CCAGTTCTCA GCCCTGGGCC TTGACCCCTG CCTTCTGGAG GACGAGCTGG 1560 ACAAGTCCGT GCAGGGCACA GGCCTGGCCT TCATCGCCTT CACTGAGGCC ATGACGCACT 1620 TCCCCACCTC CCCGTTCTGG TCCGTCATGT TCTTCTTGAT GCTTATCAAC CTGGGCCTGG 1680 GCAGCATGAT CGGGACCATG GCAGGCATCA CCACGCCCAT CATCGACACC TCCAAGGTGC 1740 CCAAGGAGAT GTTCACAGTG GGCTGCTGTG TCTTTACATT CCTCGTGGGA CTGTTGTTCG 1800 TCCAGCGCTC CGGAAACTAC TTTGTCACCA TGTTCGATGA CTACTCAGCC ACGCTGCCAC 1860 TCACTCTCAT CGTCATCCTT GAGAACATCG CTGTGGCCTG GATTTATGGA CCCAAGAAGT 1920 TCATGCAGGA GCTGACGGAG ATGCTGGGCT TCCGCCCCTA CCGCTTCTAT TTCTACATGT 1980 GGAAGTTCGT GTCTCCACTA TGCATGGCTG TGCTCACCAC AGCCAGCATC ATCCAGCTGG 2040 GGGTCACGCC CCCGGCCTAC AGCGCCTGGA TCAAGGAGGA GGCTGCCGAG CGCTACCTGT 2100 ATTTCCCCAA CTGGCCCATG GCACTCCTGA TCACCCTCAT CGTCGTGGCG ACGCTGCCCA 2160 TCCCTGTGGT GTTCGTCCTG CGGCACTTCC ACCTGCTCTC TGATGGCTCC AACACCCTCT 2220 CCGTGTCCTA CAAGAAGGCC CGCATGATGA AGGACATCTC CAACCTGGAG GAGAACGATG 2280 AGACCCGCTT CATCCTCAGC AAGGTGCCCA GTGAGGCACC TTCCCCCATG CCCACTCACC 2340 GTTCCTATCT GGGGCCCGGC AGCACATCAC CCCTGGAGAC CAGCTGGAAC CCCAATGGAC 2400 CCTATGGGCG CGGCTACCTG CTGGCCAGCA CCCCTGAGTC TGAGCTGTGA CCACTGCCCA 2460 AGCCCATGCC CGCTCTCCCC CCACCG

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 727 AMINO ACIDS
 - (B) TYPE: AMINO ACID
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Pro Lys Asn Ser Lys Val Thr Gln Arg Glu His Ser Ser Glu
5 10 15

His	Val	Thr	Glu	Ser	Val	Ala	Asp	Leu	Leu	Ala	Leu	Glu	Glu	Pro
				20					25					30
Val	Asp	Tyr	Lys	Gln	Ser	Val	Leu	Asn	Val	Ala	Gly	Glu	Ala	Gly
				35					40					45
Gly	Lys	Gln	Lys	Ala	Val	Glu	Glu	Glu	Leu	Asp	Ala	Glu	Asp	Arg
				50					55					60
Pro	Ala	Trp	Asn	Ser	Lys	Leu	Gln	Tyr	Ile	Leu	Ala	Gln	Ile	Gly
	•			65					70					75
Phe	Ser	Val	Gly	Leu	Gly	Asn	Ile	Trp	Arg	Phe	Pro	Tyr	Leu	Cys
				80		,*			85					90
Gln	Lys	Asn	Gly	Gly	Gly	Ala	Tyr	Lys	Val	Pro	Tyr	Leu	Val	Leu
				95					100					105
Leu	Ile	Ile	Ile	Gly	Ile	Pro	Leu	Phe	Phe	Leu	Glu	Leu	Ala	Val
				110					115					120
Gly	Gln	Arg	Ile	Arg	Arg	Gly	Ser	Ile	Gly	Val	Trp	His	Tyr	Ile.
				125			,		130					135
Cys	Pro	Arg	Leu	Gly	Gly	Ile	Gly	Phe	Ser	Ser	Cys	Ile	Val	
				140					145					150
Leu	Phe	Val	Gly	Leu	Tyr	Tyr	Asn	Val		Ile	Gly	Trp	Ser	
				155					160					165
Phe	Tyr	Phe	Phe	Lys	Ser	Phe	Gln	Tyr		Leu	Pro	Trp	Ser	
				170					175					180
Cys	Pro	Val	Val		Asn	Glu	Ser	Val		Val	Val	GIu	Ата	
				185				_	190	_	_	3	61	195
Cys	Glu	Lys	Ser		Ala	Thr	Thr	Tyr		Trp	Tyr	Arg	GIU	
				200	_		_		205	~ 7		T	7	210
Leu	Asp	Ile	Ser		Ser	Ile	Ser	Glu		GIĀ	GIA	ьeu	ASII	Trp
			_	215	_	_		**- 1	220	0	T1.0	<i>C</i> 1	C1.,	225 Mot
Lys	Met	Thr	Leu		Leu	Leu	vaı	Val		Ser	TTE	СТУ	СТУ	240
				230	T 1 -	71	C	Com	235	T vec	17-1	Mot	Тиг	
Ala	val	GTA	гла		тте	GTU	ser	Ser	250	пåа	val	MEC	- 7 -	255
-	a	T	Db.c	245	TT	77-7	77÷ 7	T ou		ربريو	Dhe	T.e.ii	Val	
Ser	ser	ьeu	Pne		ıyr	val	val	пец		Cys	FIIC	Leu	VUL	Arg 270
				260					265					2,0

Gly	Leu	Leu	Leu	Arg 275	Gly	Ala	Val	Asp	Gly 280	Ile	Leu	His	Met	Phe 285
Thr	Pro	Lvs	Leu		Lvs	Met	Leu	Asp		Gln	Val	Trp	Arg	
1111	110	-1-		290	-1 -				295			-	. •	300
Val	Ala	Thr	Gln	Val	Phe	Phe	Gly	Leu	Gly	Leu	Gly	Phe	Gly	Gly
				305					310					315
Val	Ile	Val	Phe	Ser	Ser	Tyr	Asn	Lys	Gln	Asp	Asn	Asn	Cys	His
				320					325					330
Phe	Asp	Gly	Ala	Leu	Val	Ser	Phe	Ile	Asn	Phe	Phe	Thr	Ser	Val
				335					340					345
Leu	Ala	Thr	Leu	Val	Val	Phe	Val	Val	Leu	Gly	Phe	Lys	Ala	
				350					355					360
Ile	Met	Asn	Glu		Cys	Val	Val	Glu		Ala	Glu	Lys	Ile	
				365					370	_	_			375
Gly	Tyr	Leu	Asn		Asn	Val	Leu	Ser		Asp	Leu	Iie	Бью	
	_			380		_	1		385	3		W	G3	390
His	Val	Asn	Phe		His	Leu	Thr	Thr		Asp	Tyr	Met	GIU	мес 405
_	Asn	**- T	T1.	395 Mat	Th.	17-1	Tira	C1.11	400	Gln.	Dhe	Ser	Δ 1 =	
Asp	ASN	vaı	TTE	Met 410	1111	Val	гуѕ	GIU	415	GIII	FIIC	361	AIG	420
C1	Leu	y an	Dro		T.611	T.011	Glu	Agn		T.e.	Asp	Lvs	Ser	
GIY	пей	Asp	PIO	425	neu	пеп	GIU	пор	430	DCG	110P	-10	002	435
Gln	Gly	Thr	Glv		Ala	Phe	Ile	Ala		Thr	Glu	Ala	Met	
	-		2	440					445					450
His	Phe	Pro	Thr	Ser	Pro	Phe	Trp	Ser	Val	Met	Phe	Phe	Leu	Met
-				455					460					465
Leu	Ile	Asn	Leu	Gly	Leu	Gly	Ser	Met	Ile	Gly	Thr	Met	Ala	${\tt Gly}_{,}$
			-	470					475					480
Ile	Thr	Thr	Pro	Ile	Ile	Asp	Thr	Ser	Lys	Val	Pro	Lys	Glu	Met
				485					490					495
Phe	Thr	Val	Gly	Cys	Cys	Val	Phe	Thr		Leu	Val	Gly	Leu	
				500					505		-1-		_	510
Phe	Val	Gln	Arg		Gly	Asn	Tyr	Phe		Thr	Met	Phe	Asp	
				515				_	520			,	a 1 -	525
Tyr	Ser	Ala	Thr	Leu	Pro	Leu	Thr	Leu	Ile	Val	Ile	Leu	Glu	Asn

				530					535					540
Ile	Ala	Val	Ala	Trp	Ile	Tyr	Gly	Pro	Lys	Lys	Phe	Met	Gln	Glu
				545				•	550			•		555
Leu	Thr	Glu	Met	Leu	Gly	Phe	Arg	Pro	Tyr	Arg	Phe	Tyr	Phe	Tyr
				560					565					570
Met	Trp	Lys	Phe	Val	Ser	Pro	Leu	Cys	Met	Ala	Val	Leu	Thr	Thr
				575					580					585
Ala	Ser	Ile	Ile	Gln	Leu	Gly	Val	Thr	Pro	Pro	Ala	Tyr	Ser	Ala
				590					595					600
Trp	Ile	Lys	Glu	Glu	Ala	Ala	Glu	Arg	Tyr	Leu	Tyr	Phe	Pro	
				605					610					615
Trp	Pro	Met	Ala	Leu	Leu	Ile	Thr	Leu	Ile	Val	Val	Ala	Thr	
				620					625					630
Pro	Ile	Pro	Val	Val	Phe	Val	Leu	Arg		Phe	His	Leu	Leu	
				635					640				. •	645
Asp	Gly	Ser	Asn		Leu	Ser	Val	Ser		Lys	Lys	Ala	Arg	•
				650					655				_	660
Met	Lys	Asp	Ile		Asn	Leu	Glu	Glu		Asp	Glu	Thr		
				665					670	_	_			675
Ile	Leu	Ser	Lys		Pro.	Ser	Glu	Ala		Ser	Pro	Met	Pro	
				680		_		_	685	0	D	T	<i>α</i> 1	690
His	Arg	Ser	Tyr		GLY	Pro	Gly	Ser		ser	Pro	Leu	GLU	
			_	695	a 3	_		01	700	G1	Ma	T 011	T 011	705
Ser	Trp	Asn	Pro		GLY	Pro	Tyr	GIY		GIA	ıyı	ьeu	ьеu	720
_	_,	_		710	GI.:	T			715					120
Ser	Thr	Pro	Glu		GIU	ьeu			·					
		-		725										

SEQUENCE LISTING

	111	GENERAL	INFORMATION
--	-----	---------	-------------

(i) APPLICANT: LI, YI

FLEISCHMANN, ROBERT

- (ii) TITLE OF INVENTION: NEUROTRANSMITTER TRANSPORTER
- (iii) NUMBER OF SEQUENCES: 6
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Human Genome Sciences, Inc.
 - (B) STREET: 9410 Key West Avenue
 - (C) CITY: Rockville
 - (D) STATE: MD
 - (E) COUNTRY: US
 - (F) ZIP: 20850
 - (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/424,424
 - (B) FILING DATE: 21-APR-1995
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Michele M. Wales
 - (B) REGISTRATION NUMBER: 43,975
 - (C) REFERENCE/DOCKET NUMBER: PF116
 - (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (301) 309-8504
 - (B) TELEFAX: (301) 309-8439
- (2) INFORMATION FOR SEQ ID NO:1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2485 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 266..2446
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGGAGGCAGG GAGTGAGGAG CGAGCGGAGT CGCGTGCGCC GGCGCGAGCT CCGGGTCGCC

CCAGCCCAG CCGGGGGCCT GTGGCGGGGG AGGAGCTGTG CGTCCGCGAC CCGTCGGGAT

-

120

CGCAG	GCTG(CT C	GGCC	GGAGI	r GC	ACGG(GCCG	AGT	CTGC	GCG A	ACTA	CCCA	CG C	GTGA	CAGGT		180
CCCTC	GAAT	GA G	AAGG	AGCTO	ACA	AGCA	GCTG	AAT	TCCA'	TCT ?	TCTC	TGTG	TG C	TGGG	GAGCA		240 ·
GGGC	raca:	CG G	CCCA	GGTG(G CA	rca i	ATG Met 1	CCG Pro	AAG Lys	AAC A	AGC Ser 5	AAA (Lys)	GTG Val	ACC Thr	CAG Gln		292
CGT (Arg (GAG Glu	CAC . His	AGC . Ser	AGT (Ser (GAG (Glu 1	CAT (GTC Val	ACT Thr	GAG Glu	TCC Ser	GTG Val	GCC Ala	GAC Asp	CTG Leu	CTG Leu 25		340
GCC (CTC Leu	GAG Glu	GAG Glu	CCT (Pro '	GTG (GAC Asp	TAT Tyr	AAG Lys	CAG Gln 35	AGT Ser	GTA Val	CTG Leu	AAT Asn	GTG Val 40	GCT Ala		388
GGT Gly	GAG Glu	GCA Ala	GGC Gly 45	GGC .	AAG Lys	CAG Gln	AAG Lys	GCG Ala 50	GTG Val	GAG Glu	GAGʻ Glu	GAG Glu	CTG Leu 55	GAT Asp	GCA Ala		436
GAG Glu	GAC Asp	CGG Arg 60	CCG Pro	GCC Ala	TGG Trp	AAC Asn	AGT Ser 65	AAG Lys	CTG Leu	CAG Gln	TAC Tyr	ATC Ile 70	CTG Leu	GCC Ala	CAG Gln		484
ATT Ile	GGC Gly 75	TTC Phe	TCT Ser	GTG Val	GGC Gly	CTC Leu 80	GGC Gly	AAC Asn	ATC Ile	TGG Trp	AGG Arg 85	TTC Phe	CCC Pro	TAC Tyr	CTG Leu		532
TGC Cys 90	CAG Gln	AAA Lys	AAT Asn	GGA Gly	GGA Gly 95	GGT Gly	GCT Ala	TAC Tyr	CTG Leu	GTG Val 100	CCC Pro	TAC Tyr	CTG Leu	GTG Val	CTG Leu 105		580
CTG Leu	ATC Ile	ATC Ile	ATC Ile	GGG Gly 110	ATC Ile	CCC Pro	CTC Leu	TTC Phe	TTC Phe 115	Leu	GAG Glu	CTG Leu	GCT Ala	GTG Val 120	GGT Gly		628
CAG Gln	AGG Arg	ATC Ile	CGC Arg 125	CGC Arg	GGA Gly	AGC Ser	ATC Ile	GGT Gly 130	Val	TGG Trp	CAC His	TAT Tyr	ATA Ile 135	Cys	CCC Pro		676
CGC Arg	CTG Leu	GGG Gly 140	Gly	ATC Ile	GGC Gly	TTC Phe	TCC Ser 145	Ser	TGC Cys	ATA Ile	GTC Val	TGT Cys 150	Leu	TTT Phe	GTG Val		724
GGG	CTG Leu 155	Tyr	TAT Tyr	AAT Asn	GTG Val	ATC Ile 160	Ile	GGG Gly	TGG Trp	AGC Ser	11e	Pne	TAT Tyr	TTC Phe	TTC Phe		772
AAG Lys 170	Ser	TTC Phe	CAG	TAC Tyr	CCG Pro 175	Leu	CCC	TGG Trp	G AGT	GAA Glu 180	ı Cys	CCT Pro	GTC Val	GTC Val	AGG Arg 185		820
AAT Asn	GG(AGC Ser	GTC Val	GCA Ala 190	. Val	GTC Val	GAC Glu	G GCA	A GAC A Glu 195	1 CAS	GAA	ı Lys	AGO Sei	TC? Sei 200	A GCC c Ala		868
ACT Thr	ACC Thi	TAC Tyi	205	e Trp	TAC Tyr	CGA	A GAG	G GCT u Ala 21	a re	G GAC u Asp	TIO	C TCT e Sei	GAG Asj 21	. se.	C ATC		916
TCG Ser	GAG	G AGC u Sei 220	r Gly	G GGC Y Gly	CTC Lev	AAC 1 Asi	TGG Tr	р ГА	G ATO	G ACC	C CT r Le	G TG(u Cy: 23(5 He	C CT	C GTG u Val	•	964

GTC Val	TGG Trp 235	AGC Ser	ATC Ile	GGG Gly	GGG Gly	ATG Met 240	GCT Ala	GTC Val	GGT Gly	AAG Lys	GGC Gly 245	ATC (CAG Ġln	TCC Ser	TCG Ser		1012
GGG Gly 250	AAG Lys	GTG Val	ATG Met	TAT Tyr	TTC Phe 255	AGC Ser	TCC Ser	CTC Leu	TTC Phe	CCC Pro 260	TAC Tyr	GTG Val	GTG Val	CTG Leu	GCC Ala 265		1060
TGC Cys	TTC Phe	CTG Leu	GTC Val	CGG Arg 270	GGG Gly	TTG Leu	TTG Leu	TTG Leu	CGA Arg 275	GGG Gly	GCA Ala	GTT Val	GAT Asp	GGC Gly 280	ATC Ile		1108
CTA Leu	CAC His	ATG Met	TTC Phe 285	Thr	CCC Pro	AAG Lys	Leu	GTC Val 290	AAG Lys	ATG Met	CTG Leu	GAC Asp	CCC Pro 295	CAG Gln	GTG Val		1156
TGG Trp	CGG Arg	GAG Glu 300	GTA Val	GCT Ala	ACC Thr	CAG Gln	GTC Val 305	TTC Phe	TTT Phe	GGC Gly	TTG Leu	GGT Gly 310	CTG Leu	GGC Gly	TTT Phe		1204
GGT Gly	GGT Gly 315	GTC Val	ATT Ile	GTC Val	TTC Phe	TCC Ser 320	Ser	TAC Tyr	AAT Asn	AAG Lys	CAG Gln 325	GAC Asp	AAC Asn	AAC Asn	TGC Cys		1252
CAC His 330	TTC Phe	GAT Asp	GGC Gly	GCC Ala	CTG Leu 335	GTG Val	TCC Ser	TTC Phe	ATC Ile	AAC Asn 340	TTC Phe	TTC Phe	ACG Thr	TCA Ser	GTG Val 345		1300
TTG Leu	GCC Ala	ACC Thr	CTC Leu	GTG Val 350	Val	TTT Phe	GTT Val	GTT Val	TTG Leu 355	GGC	TTC Phe	AAG Lys	GCC Ala	AAC Asn 360	ATC Ile		1348
ATG Met	AAT Asn	GAG Glu	AAG Lys 365	Cys	GTG Val	GTC Val	GAG Glu	AAT Asn 370	Ala	GAG Glu	AAA Lys	ATC Ile	CTA Leu 375	GGG Gly	TAC Tyr	-	1396
CTT Leu	AAC Asn	ACC Thr 380	Asn	GTC Val	CTG Leu	AGC Ser	CGG Arg 385	Asp	CTC Leu	ATC Ile	CCA	CCC Pro 390	CAC His	GTC Val	AAC Asn		1444
TTC Phe	TCC Ser	His	CTG Lev	ACC Thr	ACA Thr	AAG Lys 400	Asp	TAC Tyr	ATG Met	GAG Glu	ATG Met 405	GAC Asp	AAT Asn	GTC Val	ATC Ile		1492
ATO Met	: Thr	GTC Val	AAC L Lys	GAC Glu	G GAC 1 Asp 415	Gln	TTC Phe	TCA Ser	GCC Ala	CTG Leu 420	t GTA	CTT Leu	GAC Asp	CCC	TGC Cys 425		1540
CTT Let	r CTC	G GAC	G GAC	GA(Glu 43(r Let	GAC 1 Asp	AAG Lys	TCC Ser	GTG Val 435	. Glr	GGC Gly	ACA Thr	GGC Gly	CTG Leu 440	GCC Ala	-	1588
TTC Phe	C ATO	C GC0 € Ala	TTO a Pho 445	e Thi	r GAC r Glu	G GCC	ATO Met	ACC Thr 450	His	TTC Phe	C CCC	ACC Thr	TCC Ser 455	Pro	TTC Phe		1636
TG(Tr	TCC Se:	C GTO r Va	l Me	G TT	C TTO	C TTO	ATO 1 Met 465	t Lev	T ATO	C AAC e Asi	CTO	G GGC 1 Gly 470	r Let	GG(AGC Ser		1684
AT(G ATO	C GG e Gl	G AC	C AT	G GC	A GGC a Gly	C ATO	C ACC	C ACC	G CCC	C ATO	C ATO	GAC Asp	ACC Thi	TCC r Ser		1732

475 480 485

	4 /J																	
AAG Lys 490	GTG Val	CCC Pro	AAG Lys	GAG Glu	ATG Met 495	TTC Phe	ACA Thr	GTG Val	GGC Gly	TGC Cys 500	TGT Cys	GTC Val	TTT Phe	ACA Thr	TTC Phe 505		1780	
CTC Leu	GTG Val	GGA Gly	CTG Leu	TTG Leu 510	TTC Phe	GTC Val	CAG Gln	CGC Arg	TCC Ser 515	GGA Gly	AAC Asn	TAC Tyr	TTT Phe	GTC Val 520	ACC Thr		1828	
ATG Met	TTC Phe	GAT Asp	GAC Asp 525	TAC Tyr	TCA Ser	GCC Ala	ACG Thr	CTG Leu 530	CCA Pro	CTC Leu	ACT Thr	CTC Leu	ATC Ile 535	GTC Val	ATC Ile		1876	
CTT Leu	GAG Glu	AAC Asn 540	ATC Ile	GCT Ala	GTG Val	GCC Ala	TGG Trp 545	ATT Ile	TAT Tyr	GGA Gly	CCC Pro	AAG Lys 550	AAG Lys	TTC Phe	ATG Met		1924	
CAG Gln	GAG Glu 555	CTG Leu	ACG Thr	GAG Glu	ATG Met	CTG Leu 560	GGC Gly	TTC Phe	CGC Arg	CCC Pro	TAC Tyr 565	CGC Arg	TTC Phe	TAT Tyr	TTC Phe		1972	
TAC Tyr 570	ATG Met	TGG Trp	AAG Lys	TTC Phe	GTG Val 575	TCT Ser	CCA Pro	CTA Leu	TGC Cys	ATG Met 580	GCT Ala	GTG Val	CTC Leu	ACC Thr	ACA Thr - 585		2020	
GCC Ala	AGC Ser	ATC Ile	ATC Ile	CAG Gln 590	CTG Leu	GGG Gly	GTC Val	ACG Thr	CCC Pro 595	CCG Pro	GCC Ala	TAC Tyr	AGC Ser	GCC Ala 600	TGG Trp		2068	
ATC Ile	AAG Lys	GAG Glu	GAG Glu 605	GCT Ala	GCC Ala	GAG Glu	CGC Arg	TAC Tyr 610	Leu	TAT Tyr	TTC Phe	CCC Pro	AAC Asn 615	TGG Trp	CCC Pro		2116	
ATG Met	GCA Ala	CTC Leu 620	Leu	ATC Ile	ACC Thr	CTC Leu	ATC Ile 625	Val	GTG Val	GCG Ala	ACG	CTG Leu 630	Pro	ATC Ile	CCT		2164	
GTG Val	GTG Val 635	Phe	GTC Val	CTG Leu	CGG Arg	CAC His	Phe	CAC His	CTG Leu	CTC Leu	TCT Ser 645	Asp	GGC Gly	TCC Ser	AAC Asn		2212	
ACC Thr 650	Leu	TCC Ser	GTG Val	TCC Ser	TAC Tyr 655	Lys	AAC Lys	GCC Ala	: CGC Arg	Met 660	. Met	: Lys	GAC Asp	ATC Ile	TCC Ser 665		2260	
AAC Asn	CTG Leu	GAC Glu	GAG	Asr	Asp	GAC	ACC Thr	C CGC	TTC Phe 675	: Ile	CTC	AGC Ser	AAG Lys	GT0 Val	CCC Pro	-	2308	
AGT Ser	GAG Glu	GCA Ala	A CCT a Pro 685	Sei	CCC Pro	ATO	CCC Pro	ACT Thi	: His	CGT Arg	TCC Sei	TAT Tyr	CTC Lev 695	r GTZ	G CCC 7 Pro		2356	
GGC Gly	: AGC	700	r Sei	A CCC	CTG Lev	GAC	3 ACC 1 Thi 70!	r Sei	TGC Trp	AAC ASI	C CCC	C AAT o Asr 710	J GT7	A CCC	TAT Tyr		2404	
GGG Gly	CGC Arg 715	Gl;	C TAC y Ty:	C CTO	G CTO	G GC0 1 Ala 72	a Se	C ACC	C CC r Pro	GAC Glu	TC: 1 Se: 72	r Gli	G CTO	3 1			2446	
TGA	ACCA	CTGC	CCA	AGCC	CAT (GCCC(GCTC'	TC C	cccc	ACCG	•				,		2485	j

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 727 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
- Met Pro Lys Asn Ser Lys Val Thr Gln Arg Glu His Ser Ser Glu His
- Val Thr Glu Ser Val Ala Asp Leu Leu Ala Leu Glu Glu Pro Val Asp 20 25 30
- Tyr Lys Gln Ser Val Leu Asn Val Ala Gly Glu Ala Gly Gly Lys Gln
 35 40 45
- Lys Ala Val Glu Glu Glu Leu Asp Ala Glu Asp Arg Pro Ala Trp Asn 50 55 60
- Ser Lys Leu Gln Tyr Ile Leu Ala Gln Ile Gly Phe Ser Val Gly Leu
 65 70 75 80
- Gly Asn Ile Trp Arg Phe Pro Tyr Leu Cys Gln Lys Asn Gly Gly Gly 95
- Ala Tyr Leu Val Pro Tyr Leu Val Leu Leu Ile Ile Gly Ile Pro 100 105 110
- Leu Phe Phe Leu Glu Leu Ala Val Gly Gln Arg Ile Arg Arg Gly Ser 115 120 125
- Ile Gly Val Trp His Tyr Ile Cys Pro Arg Leu Gly Gly Ile Gly Phe 130 135 140
- Ser Ser Cys Ile Val Cys Leu Phe Val Gly Leu Tyr Tyr Asn Val Ile 145 150 155 160
- Ile Gly Trp Ser Ile Phe Tyr Phe Phe Lys Ser Phe Gln Tyr Pro Leu 165 170 175
- Pro Trp Ser Glu Cys Pro Val Val Arg Asn Gly Ser Val Ala Val Val 180 185 190
- Glu Ala Glu Cys Glu Lys Ser Ser Ala Thr Thr Tyr Phe Trp Tyr Arg 195 200 205
- Glu Ala Leu Asp Ile Ser Asp Ser Ile Ser Glu Ser Gly Gly Leu Asn 210 215 220
- Trp Lys Met Thr Leu Cys Leu Leu Val Val Trp Ser Ile Gly Gly Met 225 230 235 240
- Ala Val Gly Lys Gly Ile Gln Ser Ser Gly Lys Val Met Tyr Phe Ser
- Ser Leu Phe Pro Tyr Val Val Leu Ala Cys Phe Leu Val Arg Gly Leu 260 265 270

Leu Val Lys Met Leu Asp Pro Gln Val Trp Arg Glu Val Ala Thr Gln

Val Phe Phe Gly Leu Gly Leu Gly Phe Gly Gly Val Ile Val Phe Ser 305 310 315

Ser Tyr Asn Lys Gln Asp Asn Asn Cys His Phe Asp Gly Ala Leu Val 325 330 335

Ser Phe Ile Asn Phe Phe Thr Ser Val Leu Ala Thr Leu Val Val Phe 340 . 345 . 350

Val Val Leu Gly Phe Lys Ala Asn Ile Met Asn Glu Lys Cys Val Val 355 360 365

Glu Asn Ala Glu Lys Ile Leu Gly Tyr Leu Asn Thr Asn Val Leu Ser 370 375 380

Arg Asp Leu Ile Pro Pro His Val Asn Phe Ser His Leu Thr Thr Lys 385 390 395

Asp Tyr Met Glu Met Asp Asn Val Ile Met Thr Val Lys Glu Asp Gln 405 410 415

Phe Ser Ala Leu Gly Leu Asp Pro Cys Leu Leu Glu Asp Glu Leu Asp 420 425 430

Lys Ser Val Gln Gly Thr Gly Leu Ala Phe Ile Ala Phe Thr Glu Ala 435

Met Thr His Phe Pro Thr Ser Pro Phe Trp Ser Val Met Phe Phe Leu 450 455 460

Met Leu Ile Asn Leu Gly Leu Gly Ser Met Ile Gly Thr Met Ala Gly 465 470 475 480

Ile Thr Thr Pro Ile Ile Asp Thr Ser Lys Val Pro Lys Glu Met Phe 485 490 495

Thr Val Gly Cys Cys Val Phe Thr Phe Leu Val Gly Leu Leu Phe Val 500 505 510

Gln Arg Ser Gly Asn Tyr Phe Val Thr Met Phe Asp Asp Tyr Ser Ala 515 520 525

Thr Leu Pro Leu Thr Leu Ile Val Ile Leu Glu Asn Ile Ala Val Ala 530 535

Trp Ile Tyr Gly Pro Lys Lys Phe Met Gln Glu Leu Thr Glu Met Leu 545 550 560

Gly Phe Arg Pro Tyr Arg Phe Tyr Phe Tyr Met Trp Lys Phe Val Ser 565 570 575

Pro Leu Cys Met Ala Val Leu Thr Thr Ala Ser Ile Ile Gln Leu Gly 580 585

Val Thr Pro Pro Ala Tyr Ser Ala Trp Ile Lys Glu Glu Ala Ala Glu
595 600 605

Arg Tyr Leu Tyr Phe Pro Asn Trp Pro Met Ala Leu Leu Ile Thr Leu Ile Thr Leu Wal Val Val Ala Thr Leu 630 Pro Ile Pro Val Val Phe Val Leu Arg His 640 Phe His Leu Leu Ser Asp Gly Ser Asn Thr Leu Ser Val Ser Tyr Lys Ala Arg Met Met Lys Asp Ile Ser Asn Leu Glu Glu Asn Asp Glu Thr Arg Phe Ile Leu Ser Lys Val Pro Ser Glu Ala Pro Ser Pro Met

Pro Thr His Arg Ser Tyr Leu Gly Pro Gly Ser Thr Ser Pro Leu Glu

695

Thr Ser Trp Asn Pro Asn Gly Pro Tyr Gly Arg Gly Tyr Leu Leu Ala

Ser Thr Pro Glu Ser Glu Leu 725

- (2) INFORMATION FOR SEQ ID NO:3:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GACTAAAGCT TGGCATCAAT GCCGAAGAAC

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAACTTCTAG AGCAGTGGTC ACAGCTCAG

- (2) INFORMATION FOR SEQ ID NO:5:
 - (i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 38 base pairs

30

29

- 9
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
 GACTAAGATC TGCCACCATG CCGAAGAACA GCAAAGTG

20

- (2) INFORMATION FOR SEQ ID NO:6:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
 GAACTGATAT CGCAGTGGTC ACAGCTCAG

29